Abnormality Detection in Electrocardiograms by Time Series Alignment

Bachir Boucheham

Department of Informatics, University of Skikda 20 Août 1955
BP 26, Route El-Hadaek, Skikda, DZ21000 Algeria

boucbac@yahoo.com

***Abstract*-Electrocardiogram (ECG) morphology deviation from normal beat is a sign of abnormal behavior and a significant information for cardiologist to depict cardiac diseases. Most existing methods for such a task use single beat classification with various tools. However, these approaches usually ignore the repetitive nature of the ECG. In this study, we adapt and apply a previously developed method for quasi-periodic time series comparison (SEA) to detect the morphology change in the ECG. The basic idea is to perform segment-wise comparisons of the ECG where one segment stands for the reference (normal) behavior and the other segment for the unknown behavior segment. Due to many difficulties, this is a very complex problem to solve, especially with regard to the phase shift and the number of periods in each segment problems. The new approach is applied on records from the Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH) arrhythmia database. Results show the effectiveness of the proposed method in detecting the significant morphology changes of the ECG. The author believes that the method could also be useful for clustering and summarizing of ECG, among other applications.**

Keywords-Electrocardiograms (ECG); Anomaly Detection; Pattern Recognition; Time Series Comparison; Shape Exchange Algorithm (SEA)

I. INTRODUCTION

The electrocardiogram (ECG) is a signal that reflects heart activity. The ECG is widely used by medical doctors for depicting any heart dysfunctions. The ECG itself is a concatenation of cycles (pseudo-periods) that are normally quasi-similar in shape, where each cycle corresponds to one heartbeat. Each period is mainly composed of three complexes in this order: P-wave, QRS complex and T-wave. Figure 1 illustrates an ECG segment with its main components. The ECG is globally classified by medical experts as *normal* or *abnormal* (heartbeat). The automation of this last task is of major importance for long ECG analysis and diagnosis and for patient surveillance. The main sign of the transition of the ECG from normal to abnormal is a significant change in the ECG *morphology*. Figure 2 illustrates one case of this transition. The figure shows an ECG trace where the first seven cycles are normal, whereas the eighth beat is abnormal (Premature Ventricular Contraction, PVC). As can be seen, the main characteristic of this last beat is its specific morphology with regard to the previous (normal) beats. Many methods have been proposed for ECG pattern discrimination, including abnormal heartbeats detection. These approaches range from frequency domain analysis e.g. [1] to neural networks e.g. [2], discriminant analysis e.g. [3], SVM e.g. [4], self organizing maps e.g. [5], Gaussian mesa function models and nonlinear probability estimators [6], Hermite Functions combined to Self-Organizing Maps [7], neuro-fuzzy network [8], higher order - 6 -statistics [9], grey relational analysis-based classifier [10], etc.

However, these methods ignore the basic nature of the ECG of being a quasi-periodic signal and proceed to a single-beat analysis, which assumes ECG segmentation in basic cycles, mostly through QRS detection. In a previous work, we designed an algorithm (SEA) that is able to match quasi-periodic time series with no specific mandatory segmentation [11]. SEA is based on a pattern recognition approach through time series comparison and alignment. As a time series comparator and aligning tool, SEA is distinguished by its unique characteristic of being able to match two (phase-shifted) time series of arbitrary lengths and containing an arbitrary number of periods each. To the author's best knowledge, this kind of comparison and alignment are the most complex time series alignments that only SEA can perform. For instance, in the case of the ECG, SEA exploits the quasi-repetitive nature of this signal to derive a similarity measure that can tell if two ECGs are similar or not. More, SEA is able to align ECGs on a point-to-point level. This property enables SEA to yield a kind of difference between the two ECG traces. This is important for localizing the main differences between the two ECGs. In the current study, these abilities are exploited for the detection of ECG morphology change. The method mainly consists in considering a *reference* normal (known behavior) ECG segment R composed of few (say, 3 to 5) beats and scan the input ECG (unknown behavior) segment-wise. Each *segment* of the input ECG T is compared to the *reference* segment using SEA. The SEA method outputs a measure in the range [0..1] in terms of the correlation factor (*Corr*, Eq. 1) that reflects the degree of similarity between the reference and the segment under analysis. Based on this measure, the whole segment is classified as *normal* or *abnormal*. It should be noted that in fact, the measure is performed on the segment T and its reconstructed T_{Rec} by SEA, as will be explained below. We also stress the point that we do not perform a cycle-by-cycle classification. We rather class the entire segment T as abnormal or normal depending on the occurrence or not of any detected change in the cycles morphology with respect to the reference R. Therefore, and to all evidence, no specific segmentation of the ECG in basic cycles is mandatory. The method was tested on selected ECG records from the Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH) ECG database. Results show the effectiveness of the approach in detecting and localizing morphological changes of the ECG corresponding to occurrences of different types of abnormal heartbeats.

$$corr(\mathrm{T},\mathrm{T}_{\mathrm{Rec}}) \frac{\mathrm{cov}(T,T_{\mathrm{Rec}})^2}{\mathrm{var}(T).\mathrm{var}(T_{\mathrm{Rec}})} \quad (1)$$

The rest of this study is organized as follows. In section II, materials and methods in relation with this study are presented. In section III, applications using the proposed method are reported with obtained results. In section IV, obtained results and the method are discussed. At last, in section V, this work is concluded and future plans are announced.

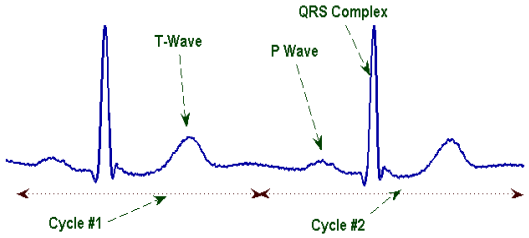


Fig. 1. A Typical Normal ECG Segment Composed of Two Cycles

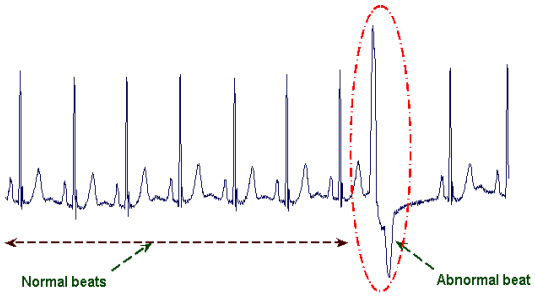


Fig. 2. An ECG Segment Illustrating the Morphology Change from Normal (first 7 beats) to Abnormal (Beat #8, Premature Ventricular Contraction, PVC)

II. MATERIA` 1LS AND METHODS

In this section, we present a brief description on time series comparison in general, followed by the proposed technique.

A. Time Series Comparison

Briefly, there are two classes of time series matching techniques: Comparison methods and alignment methods. Let R=(r_i), i=1:n and T=(t_j), j=1:m, be two given time series. The first class renders a distance that reflects the degree of similarity between R and T; whereas alignment methods perform a mapping between the points of R and those of T. Many methods have been proposed in the literature for this problem, including the Discrete Fourier Transform (DFT) e. g. [12], string matching algorithms e.g. [13], histogram comparison e.g. [14], Edit distance e.g. [15] and the dynamic time warping (DTW) method e.g. [16]. This last method is recognized by many authors as the most accurate alignment technique.

Particularly, the Shape Exchange Algorithm (SEA) [11] is a method that was recently proposed for time series matching. Its superiority over the popular DTW technique was demonstrated. Due to its many nice properties, the SEA method is adapted and used in this study as a tool for detection of morphology change in ECGs. For a thorough description of that method, the reader is invited to consult reference [11]. The proposed algorithm for detection of morphology change in ECGs is described in the next sub-section.

B. The Proposed Algorithm

Let *R* be the known normal heartbeat trace (length *n*); *E* be the input ECG (length potentially infinity); *T* be a segment from *E* (length *m*); then, the proposed algorithm consists in the following steps:

1) Acquisition of R and E ECG Patterns: In this step, *R* and *E* are fed to the algorithm. These can be specified according to different ways: read from a stocked file or specified by onset-offset settings on a record, etc.

2) Scanning: The Input ECG *E* is scanned left to right. In each step, a segment *T* of length *m* is extracted from the current position. For each segment *T*, apply steps *3)-8)* bellow.

3) Sorting: In this step, the time series to compare (*R* and *T*) are sorted on the magnitude value. This yields for each series two indexes: The sorted magnitudes index and the temporal index indicating for each value its time occurrence.

4) Magnitude Exchange: In this step, the sorted magnitudes are exchanged between the two time series. That is *T* will get the sorted magnitudes of *R,* and vice-versa.

5) Reconstruction: In this step, there is reconstruction of the two time series resulting from the sorted magnitudes exchange. This is done by *re-ordering* the two time series (magnitudes and time indexes) on the temporal indexes. The result of this operation for *T* is a new time series T_{Rec} corresponding to the reconstructed time series of *T* using time series *R* magnitudes.

6) Comparison: In this step, the correlation factor (Eq. 1) is computed for the couple (T,T_{Rec}). The closer this value to 1, the more the two corresponding time series are similar.

7) Special Step: In case the two time series are of different lengths, a linear interpolation is performed between the sorted magnitudes of *R* and those of *T,* prior to the exchange step *4)*.

8) Decision: A user specified threshold value on the correlation factor *c* is used for decision as follows:

If Corr(T,TRec) < c then 'There-is-change'

Else 'no-change-detected'

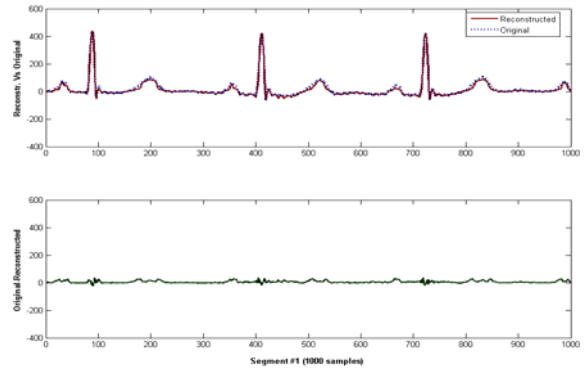
III. EXPERIMENTAL RESULTS

In this section, applications of the proposed algorithm are presented. We first stress the need to set the threshold value *c* of step *4)* (Decision) in the proposed algorithm. Through intensive experiments, we found that correlation factors that are greater than 0.950 indicate strong resemblance between the compared segments. Therefore, in the step above, *c*=0.950. We also mention that all used segments are selected from the MIT-BIH ECG database, which is sampled at 360 Hz.

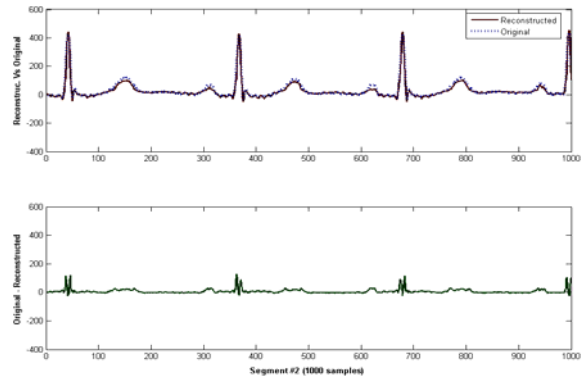
A. Application #1

This is an illustrative application of the method. The used segments are as follows. The reference segment R is that in Fig. 1. It is extracted from the beginning of rec. MITBIH#103

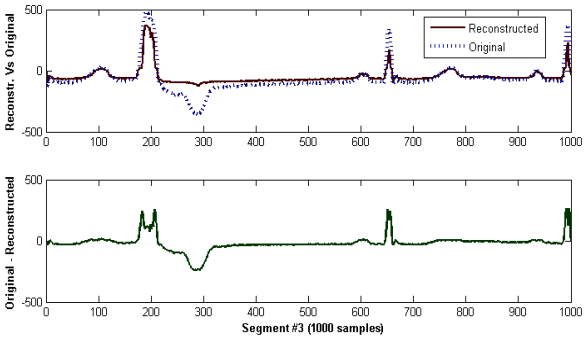
and is n=600 samples long. The input ECG is that in Fig. 2. It is extracted from rec.MITBIH#119 and is 3000 samples long. Note that, even though the two ECGs are extracted from two different records, they are similar (by visual judgment) in some regions. In order to locate any abnormalities, a window T of length m=1000 samples is used. This requires three comparisons to perform. They are presented in Fig.3 a-c. All figures are plotted in the same fashion. The upper subplot shows the current window (T: discontinuous line) versus the reconstructed window upon application of the SEA method (T_{Rec} continuous line). The lower subplot shows the difference between the original window (T) and T_{Rec}. The obtained correlations and decisions for the three windows are reported in Table 1.



a)



b)



c)

Fig. 3. An Illustrative Example of the Used Method: (a-b) No Morphology Change (Normal Beats), (c) Morphology Change (PVC Occurrence, 1 Abnormal Beat)

TABLE I
OBTAINED RESULTS FOR APPLICATION #1

Window#	Correlation	Decision
1	0.992	Normal
2	0.978	Normal
3	0.916	Abnormal

TABLE II
OBTAINED RESULTS FOR APPLICATION #2, (FIG.4)

Window #	Correlation	Morphology Change Detection?
1	0.992	No
2	0.978	No
3	0.916	***Yes***
4	0.992	*No*
5	0.918	***Yes***
6	0.859	***Yes***
7	0.996	No
8	0.921	***Yes***
9	0.910	***Yes***
10	0.992	No

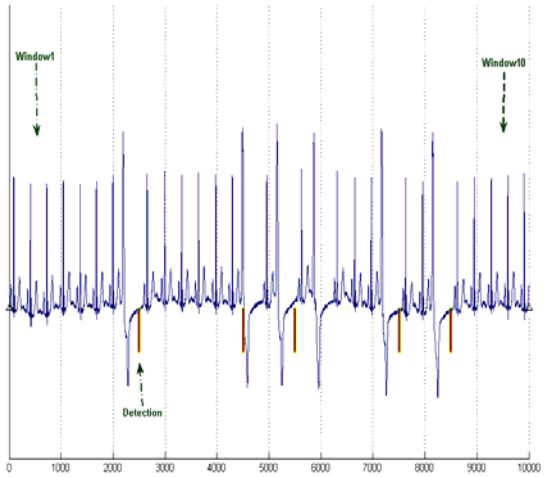


Fig. 4. Detection of morphology change (windows 3,5,6,8 and 9), in this case PVCs, within record MITBIH #119. The windows are compared with the normal ECG morphology reported in Fig.1

B. Application #2

In this application, we further show the effectiveness of the proposed method on long ECGs. The second experiment is performed using as reference R the segment in Fig.1 and the same input ECG as in the first application. However, for the input ECG, we use a longer segment of 10000 samples. This input is shown in Fig.4. This figure reports also the windows (T) delineations (vertical discontinuous lines). The methodology for this experiment is to follow the input ECG morphology change with respect to R and the ability of the proposed method in detecting such events. The long thick bars, centred at some windows, indicate a 'morphology change

detection'. As can be seen, the proposed method did detect all the significant morphology changes, which in this application all correspond to Premature Ventricular Contraction (PVC). The correlation factors corresponding to the ten windows are reported in Table 2.

IV. DISCUSSION

The applications clearly show that when the compared segments are similar (visual judgment), the correlation factor is well above the threshold value c=0.950. Results show also that in this case, the reconstructed window T_{Rec} is almost undistinguishable from the original window T. Also, which is an obvious consequence of the above results, the difference between T and T_{Rec} is significantly small (e.g. Fig.3a-b). On the contrary, when there is a significant change in the morphology of the current window T with comparison to the reference segment R, results are all the way around. The correlation factor is significantly below the c value and the plots show significant differences in T_{Rec} with comparison to T. These results are confirmed by the difference $T - T_{Rec}$ plot, which indicates in these cases the locations of discord between T and T_{Rec} (e.g. Fig.3c). This is in fact an important result, in the sense that not only the method can tell if there is change in the ECG morphology, but it also can localize the occurring time of this event. Note also that the method is capable of comparing phase shifted segments of different lengths and/ or containing an arbitrary number of periods each (e.g. Fig.1 Versus Fig.2.a). This grants it with the flexibility of requiring almost no conditions on the compared ECGs, including the segmentation lead by QRS detection. For example, in application 1, R is of length 600 samples and contains two periods, whereas the three windows in Fig. a-c are all of length 1000 samples and contain respectively 3, 4 and 3 periods. Another characteristic of the method is its property of being nearly parameter-less. Indeed, except the c value and the lengths of the reference R (n) and the window T (m), which practically are not difficult to set, no other parameter is required. In this context and as a comparison case, Jekova et al [3] used twenty six extracted parameters from pre-processed ECG records for premature ventricular contraction (PVC) and Normal (N) beat classification. The discrimination itself was performed through discriminant analysis of the extracted parameters, step that itself extracts another great number of parameters. Another specificity of the method should be reported. As can be seen on Table 2, the value of the correlation factor of window 6 is significantly lower than the values of the other windows with a detection (3,5,8 and 9). By having a look at Fig.4, it can be checked that window 6 contains in fact two PVCs, whereas the other windows (3,5,8 and 9) contain one PVC only. This is quite an interesting indicator of the degree of morphology change in the considered window.

To all evidence, the normal ECG morphology is not restricted to the case of Fig.1. But, this is the most encountered normal ECG pattern. Also, the morphology change is not restricted to the case of the PVC. However, this anomaly was considered in this study as an illustration case for the proposed method. We conducted intensive experiments with other types of normal ECGs as reference which were compared to ECGs containing different types of morphology change and the results confirmed the effectiveness of the proposed method.

V. CONCLUSION

In this study, we proposed a new method for the detection of the morphology change in ECGs. The proposed method is based on a pattern recognition approach through usage of a time series comparison and alignment technique (SEA). The method effectiveness was demonstrated through many examples. Specifically, the method was shown to be effective in detecting the morphology change within ECGs with high precision, based on the correlation factor between the considered window and its reconstructed correspondent upon application of the SEA method. The method is characterized by several interesting properties, including its ability to compare ECG segments with different lengths, eventually phase shifted and containing an arbitrary number of periods each. The proposed method can be used in fact for many purposes, including anomaly detection in ECGs, on which the illustrations of this study were centred, ECG patterns classification and Clustering. For future works, these and other applications will be investigated. Integration of the method in specialized software for automatic patient surveillance or automatic ECG diagnosis are also interesting issues.

ACKNOWLEDGEMENT

This work is partially supported by the Ministry of Higher Education and Scientific Research of the Algerian Government through a MESRS-CNEPRU Grant.

REFERENCES

[1] C-H Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier", Computers and Mathematics with Applications, Vol. 55, pp. 680–690, 2008.

[2] W. G. Baxt, "Application of artificial neural networks to clinical medicine", Lancet, pp. 1135–1138, 1995.

[3] I Jekova, G Bortolan, and I Christov, "Pattern Recognition and Optimal Parameter Selection", in Premature Ventricular Contraction Classification, Computers in Cardiology, Vol. 31, pp. 357−360, 2004.

[4] A.H. Khandoker, Y. Kimura and M Palaniswami, "Automated Identification of Abnormal Fetuses Using Fetal ECG and Doppler Ultrasound Signals", Computers in Cardiology, Vol. 36, pp. 709−712, 2009.

[5] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps", IEEE Tans. Biomed. Engin., Vol. 47, no. 7, 2000.

[6] R. Duboisa, P. Maison-Blancheb, B. Queneta, and G. Dreyfusa, "Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators", Computer Methods and Programs in Biomedicine, Elsevier, Vol. 88, pp. 217–233, 2007.

[7] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and Leif Sörnmo, "Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps", IEEE Trans. On Biomed. Engin., Vol. 47, no 7, 2000.

[8] M. Engin, ECG beat classification using neuro-fuzzy network, Pattern Recognition Letters, Elsevier, Vol. 25, pp. 1715–1722, 2004.

[9] S-N. Yu, Y-H. Chen, "Noise-tolerant electrocardiogram beat classification based on higher order statistics of subband components", Artificial Intelligence in Medicine, Elsevier, Vol. 46, pp. 165-178, 2009.

[10] C-H. Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier", Computers and Mathematics with Applications, Vol. 55, pp. 680–690, 2008.

[11] B. Boucheham, "Matching of quasi-periodic time series patterns by exchange of block-sorting signatures", Pattern Recognition Letters, Elsevier, Vol. 29, pp. 501-514, 2008.

[12] R. Agrawal, C. Faloutsos, and A.N. Swami, "Efficient similarity search in sequence databases", In: Proc. 4th Int. Conf. of Foundations of Data Organization and Algorithms", 69-84, 1993.

[13] H. Shatkay and S.B. Zdonik, "Approximate queries and representations for large data sequences", In: Proc. 4th Intern. Conf. on Data Engin., pp. 536–545, 1996.

[14] L. Chen and M.T. Özsu, "Similarity based retrieval of time series data using multi-scale histograms", CS Tech. Report, CS-2003-31, School of Computer Science, University of Waterloo, 2003.

[15] T. Bozkaya, N. Yazdani, and Z.M. Ozsoyoglu, "Matching and indexing sequences of different lengths", In Proc. 6th Int. Conf. on Information and Knowledge Management, pp. 128-135, 1997.

[16] V. Tuzcu and S. Nas, "Dynamic time warping as a novel tool in pattern recognition of ECG changes in Heart Rhythm disturbances", IEEE Intern. Conf. On Systems, Man and Cybernetics, Waikoloa, Hawai, October 10-12, pp. 182-185, 2005.